BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e35216
doi: 10.3897/biss.3.35216

OPEN ACCESS

Conference Abstract

Using Crowd-curation to Improve Taxon Annotations on the Wikimedia Infrastructure

Andra Waagmeester[‡], Daniel Mietchen[§], Siobhan Leachman[|], Quentin Groom[¶]

‡ Micelio, Ekeren (Antwerpen), Belgium
§ University of Virginia, Charlottesville, United States of America
| Citizen Scientist, Wellington & Wairarapa, New Zealand
¶ Meise Botanic Garden, Meise, Belgium

Corresponding author: Andra Waagmeester (andra@micelio.be)

Received: 05 Apr 2019 | Published: 18 Jun 2019

Citation: Waagmeester A, Mietchen D, Leachman S, Groom Q (2019) Using Crowd-curation to Improve Taxon Annotations on the Wikimedia Infrastructure. Biodiversity Information Science and Standards 3: e35216. https://doi.org/10.3897/biss.3.35216

Abstract

Here we present how two independent infrastructures, Wikimedia and iNaturalist, can be jointly leveraged to improve content on both platforms. iNaturalist.org began as a Master's final project in 2008 and grew to a globally used app to help identify biodiversity. The community behind iNaturalist consists of citizen scientists, who record a species existence through photos or sound recordings. The Wikimedia Foundation provides a spectrum of resources, of which Wikipedia is the most famous sibling. Other siblings we address here are Wikimedia Commons and Wikidata. Commons is the platform where open licensed media can be shared. Wikidata is the linked knowledge graph, where public data can be stored as structured data. Basically, data goes to Wikidata, images and recordings go to Commons and text goes to Wikipedia. Initially, both Commons and Wikidata served primarily the approximately 300 language versions of Wikipedia. However, nowadays both Commons and Wikidata are also being used as is, that is in other contexts than Wikipedia.

Although iNaturalist and the Wikimedia family of repositories thrive as independent infrastructures and are thus maintained independently from each other, they can be mutually beneficial to each other. Content created by the iNaturalist community can be very valuable to the Wikimedia community.

© Waagmeester A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Firstly, observations in the form of photos and sound recordings can be stored on iNaturalist, with an open and compatible license that can provide valuable illustrations and structured knowledge on biodiversity. Wikipedia articles can be enriched with already approximately 1.2M photos from observations. Furthermore, the assertions made by the iNaturalist community can act as references in various Wikidata claims or Wikipedia articles. In many cases in Wikipedia, we have to rely on personal annotations by the picture taker, who stores it on Wikimedia's multimedia Commons. iNaturalist provides images of organisms with stronger - peer-reviewed - assertions on the subject in the picture. When a cat is called a cat in Wikipedia or Commons, it is the iNaturalist community that either approves or rejects that claim.

In the opposite direction, iNaturalist relies on knowledge described in Wikipedia. It includes Wikipedia articles about taxa on its website. Images can also be uploaded into iNaturalist from Commons. It is therefore possible to add images of museum specimens into iNaturalist to assist with taxon identification.

Other citizen science apps do exist. iNaturalist, however, is particularly interesting due to the feature that it allows its users to select from a selection of licenses, of which some are compatible with the licenses upon which content from the Wikimedia family is available. Wikidata uses a CC0 - public - license for its data, Wikipedia is available under a CC-BY-SA and Commons content uses a selection of Creative Commons licenses (Table 1). iNaturalist is also a particular good fit with Wikimedia, because both have a global and multilingual scope. This is a great example of how platforms can support each other's missions by simple policy decisions, such as open licencing, that underpin interoperability.

Table 1.

Availability of licenses on the different platforms of Wikimedia resources and iNaturalist.

Licenses available on iNaturalist	Add to iNaturalist	Add to Wikidata	Add to Commons	Add to Wikipedia
CC0	+	+	+	+
CC-BY	+	+	+	+
CC-BY-NC	+	-	-	-
CC-BY-SA	+	-	+	+
CC-BY-ND	+	-	-	-
CC-BY-NC-SA	+	-	-	-
CC-BY-NC-ND	+	-	-	-

Keywords

crowd curation, iNaturalist, Wikidata, Wikimedia, Creative Commons

Presenting author

Andra Waagmeester